

# On discrimination between two close distribution tails.

Rodionov I. V.\*

February 21, 2017

## 1 Introduction. Main result.

Statistics deals often with discrimination of close distributions based on censored or truncated data, in particular, for high-risk insurances and reliability problems. The situation when one observes data exceeding a pre-determined threshold is well-studied, see [1], [2], [3] and references therein. On the other hand statistics of extremes says that only higher order statistics should be used for discrimination of close distribution tails, wherein moderate sample values can be modeled with standard statistical tools. In particular, such approach for distributions from Gumbel maximum domain of attraction (for the definitions see [4]) is considered in [5], [6], [7]. As well, any estimators of the extreme value indices  $\gamma$  and  $\rho$  (see [8]) can be used also to discriminate the distribution tails. Notice that we do not assume belonging the corresponding distribution function to a maximum domain of attraction.

**Definition 1** *The distribution functions  $F$  and  $G$  are said to be satisfied the condition  $B(F, G)$  if for some  $\varepsilon > 0$  and  $x_0$*

$$\frac{1 - G(x)}{(1 - F(x))^{1-\varepsilon}} \text{ is nondecreasing with } x > x_0. \quad (1)$$

Denote by  $\Theta(F_0)$  the class of continuous distribution functions  $F_1$  satisfying either  $B(F_1, F_0)$  or  $B(F_0, F_1)$ . Consider the simple hypothesis  $H_0 : F = F_0$  and the alternative hypothesis  $H_1 : F \in \Theta(F_0)$ , where  $F_0$  is continuous. Notice that if distribution functions  $F, G$  satisfy either  $B(F, G)$  or  $B(G, F)$  for some  $\varepsilon > 0$  then it holds for all  $\varepsilon_1$ ,  $0 < \varepsilon_1 < \varepsilon$ . So denote

$$\varepsilon(F, G) = \max\{\varepsilon : F, G \text{ satisfy either } B(F, G) \text{ or } B(G, F) \text{ for } \varepsilon\}.$$

---

\*Moscow State University, Faculty of Mathematics and Mechanics and Moscow Institute of Physics and Technology, Faculty of Innovations and High Technologies. E-mail: vecsell@gmail.com

Denote by  $\Theta_\varepsilon(F_0)$  the class of continuous distribution functions  $F_1$  satisfying either  $B(F_1, F_0)$  or  $B(F_0, F_1)$  with  $\varepsilon(F_0, F_1) \geq \varepsilon$  and consider another alternative hypothesis  $H_{1,\varepsilon} : F \in \Theta_\varepsilon(F_0)$ . Let  $X_1, \dots, X_n$  be i.i.d. random variables with a common distribution function  $F$ . Denote by  $X_{(1)} \leq \dots \leq X_{(n)}$  the order statistics for them. Introduce the Hill-like statistics

$$R_{k,n} = \ln(1 - F_0(X_{(n-k)})) - \frac{1}{k} \sum_{i=n-k+1}^n \ln(1 - F_0(X_{(i)})).$$

which we are going to use for the problem of discrimination between the two introduced above hypotheses when  $k$  higher order statistics are known. Remark that if  $F_0$  is Pareto distribution function with parameter  $\gamma$ , then

$$R_{k,n} \stackrel{d}{=} \gamma_H / \gamma,$$

where  $\gamma_H$  is the Hill estimator of  $\gamma$ . If furthermore  $F_0$  belongs to Fréchet max-domain of attraction, then  $R_{k,n}$  behaves asymptotically as  $\gamma_H / \gamma$ , that is, their ratio tends to one as  $n \rightarrow \infty$ . We will show that the distributions of  $R_{k,n}$  if either  $H_0$  or  $H_1$  fulfilled are different which can give a statistical for discrimination the hypotheses. The following two results describe the behavior of  $R_{k,n}$  as  $k, n \rightarrow \infty$  with  $k < n$  provided  $H_0$  or  $H_1$  is fulfilled.

**Theorem 1** *If  $H_0$  holds then*

$$\sqrt{k}(R_{k,n} - 1) \xrightarrow{d} \xi \text{ as } k, n \rightarrow \infty,$$

where  $\xi$  is standard normal random variable, i.e.  $\xi \sim N(0, 1)$ .

This theorem gives obvious goodness-of-fit test for the tail of  $F$ . Besides, the following result provides some information about the consistency of this test. Assume that  $H_0$  does not hold and  $F$  is equal to  $F_1$  which is different from  $F_0$ . Denote  $x^*$ , the right endpoint of  $F_1$ , that is,  $x^* = \sup\{x : F_1(x) < 1\}$ . Assume that  $F_0$  and any  $F_1 \in \Theta(F_0)$  have the same right endpoint (how to discriminate distributions with different endpoints, see [10], [4]). Further consider  $x^* = +\infty$ , otherwise change variables  $y = 1/(x^* - x)$  gives the assumption. The following theorem shows consistency of the proposed test.

**Theorem 2** (i) *If  $H_1$  holds then*

$$\sqrt{k_n}|R_{k_n,n} - 1| \xrightarrow{d} +\infty$$

provided  $k_n \rightarrow \infty$ ,  $k_n/n \rightarrow 0$  as  $n \rightarrow \infty$ .

(ii) *If  $H_{1,\varepsilon}$  holds then under the same conditions*

$$\inf_{F_1 \in \Theta_\varepsilon(F_0)} \sqrt{k_n}|R_{k_n,n} - 1| \xrightarrow{d} +\infty.$$

The considered test makes it possible to discriminate, for example, two normal distributions with different variances, but we should weaken the condition (1) to discriminate two normal distributions with the same variance and different means. But weakening the condition (1) imposes some conditions on behavior of the sequence  $k_n$ .

**Definition 2** *The distribution functions  $F$  and  $G$  are said to satisfy the condition  $C(F, G)$  if for some  $\varepsilon > 0$  and  $x_0$*

$$\frac{1 - G(x)}{(1 - F(x))(-\ln(1 - F(x)))^\varepsilon} \text{ is nondecreasing, } x > x_0. \quad (2)$$

Denote by  $\Theta'(F_0)$  the class of continuous distribution functions  $F_1$  satisfying either  $C(F_1, F_0)$  or  $C(F_0, F_1)$  and the following condition: for some  $\delta \in (0, 1)$

$$1 - F_1(x) \leq (1 - F_0(x))^\delta, \quad x > x_0. \quad (3)$$

See, if distribution functions  $F, G$  satisfy either  $C(F, G)$  or  $C(G, F)$  for some  $\varepsilon > 0$  then it holds for all  $\varepsilon_1$ ,  $0 < \varepsilon_1 < \varepsilon$ . Denote

$$\varepsilon'(F, G) = \max\{\varepsilon : F, G \text{ satisfy either } C(F, G) \text{ or } C(G, F) \text{ for } \varepsilon\}.$$

Denote by  $\Theta'_\varepsilon(F_0)$  the class of continuous distribution functions  $F_1$  satisfying (3) and either  $C(F_1, F_0)$  or  $C(F_0, F_1)$  with  $\varepsilon'(F_0, F_1) \geq \varepsilon$ . As before, consider the simple hypothesis  $H_0 : F = F_0$  and two alternative hypotheses  $H'_1 : F \in \Theta'(F_0)$ ,  $H'_{1,\varepsilon} : F \in \Theta'_\varepsilon(F_0)$  with continuous  $F_0$ .

**Theorem 3** (i) *If  $H'_1$  holds then*

$$\sqrt{k_n}|R_{k_n,n} - 1| \xrightarrow{d} +\infty$$

*provided  $k_n/n \rightarrow 0$ ,  $k_n^{1/2-\alpha}/\ln n \rightarrow +\infty$ , for some  $\alpha \in (0, 1/2)$ , as  $n \rightarrow \infty$ .*

(ii) *If  $H'_{1,\varepsilon}$  holds then under the same conditions*

$$\inf_{F_1 \in \Theta'_\varepsilon(F_0)} \sqrt{k_n}|R_{k_n,n} - 1| \xrightarrow{d} +\infty.$$

## 2 Auxiliary results and proofs.

### 2.1 Auxiliary results.

Since  $R_n$  depends on the higher order statistics we cannot immediately use independence of the random variables  $(X_1, \dots, X_n)$ . Therefore consider the conditional distribution of  $R_n$  given  $X_{(n-k)} = q$  applying the following lemma.

**Lemma 1** ([4]) *Let  $X, X_1, \dots, X_n$  be i.i.d. random variables with common distribution function  $F$ , and let  $X_{(1)} \leq \dots \leq X_{(n)}$  be the  $n$ th order statistics. For any  $k = 1 \dots n-1$ , the conditional joint distribution of  $\{X_{(i)}\}_{i=n-k+1}^n$  given  $X_{(n-k)} = q$  is equal to the (unconditional) joint distribution of the corresponding set  $\{X_{(i)}^*\}_{i=1}^k$  of order statistics for i.i.d. random variables  $\{X_i^*\}_{i=1}^k$  having the distribution function*

$$F_q(x) = P(X \leq x | X > q) = \frac{F(x) - F(q)}{1 - F(q)}, \quad x > q.$$

We call  $F_q(x)$ ,  $x > q$ , the tail distribution function linked with the distribution function  $F$ . Consider two continuous distribution functions  $F$  and  $G$  and a random variable  $\xi_q$  with distribution function  $G_q$ , where  $q \in \mathbb{R}$  is some parameter. Let

$$\eta_q = \ln \left( \frac{1 - F(q)}{1 - F(\xi_q)} \right).$$

Clear,  $\eta_q \geq 0$  for all  $q \in \mathbb{R}$ .

The crucial point in the proof of Theorem 2 is studying of asymptotical behavior of  $\eta_q$ .

**Proposition 1** *Let  $F_q$  and  $G_q$  are tail distribution functions of  $F$  and  $G$  respectively. Then*

- (i) *If for some  $x_0$ ,  $q > x_0$ , and any  $x > q$ ,  $F_q(x) = G_q(x)$ , then  $\eta_q$  is standard exponential.*
- (ii)  *$G_q(x) \geq F_q(x)$  for any  $x > q$  if and only if  $\eta_q$  is stochastically smaller than a standard exponential random variable.  
 $G_q(x) \leq F_q(x)$  for any  $x > q$  if and only if  $\eta_q$  is stochastically larger than a standard exponential random variable.*
- (iii)  *$G_q(x) \geq F_q(x)$  for any  $x > q \geq x_0$  and some  $x_0$  if and only if  $(1 - G(x))/(1 - F(x))$  is nonincreasing function as  $x > x_0$ .*

## 2.2 Proof of Proposition 1.

(i) Let  $F_q(x) = G_q(x)$  for all  $x > q$ , then we have for the distribution function of  $\eta_q$ ,

$$\begin{aligned} P(\eta_q \leq y) &= P\left(\ln \left( \frac{1 - F(q)}{1 - F(\xi_q)} \right) \leq y\right) = P\left(\frac{1 - F(q)}{1 - F(\xi_q)} \leq e^y\right) = \\ &= P\left(F(\xi_q) \leq 1 - (1 - F(q))e^{-y}\right) = P\left(\xi_q \leq F^{\leftarrow}\left(1 - \frac{1 - F(q)}{e^y}\right)\right). \end{aligned} \quad (4)$$

Furthermore, for the same  $x$ ,

$$P\left(\xi_q \leq F^{\leftarrow}\left(1 - \frac{1 - F(q)}{e^y}\right)\right) = \frac{F\left(F^{\leftarrow}\left(1 - \frac{1 - F(q)}{e^y}\right)\right) - F(q)}{1 - F(q)} = 1 - e^{-y}.$$

(ii) Now assume that for all  $x > q$  and some  $q \in \mathbb{R}$ ,  $G_q(x) \geq F_q(x)$ . Then from (4), since  $1 - (1 - F(q))e^{-y} \geq F(q)$  for all  $y \geq 0$  it follows that

$$P(\eta_q \leq y) = \frac{G\left(F^{\leftarrow}\left(1 - \frac{1 - F(q)}{e^y}\right)\right) - G(q)}{1 - G(q)} \geq$$

$$\frac{F\left(F^{\leftarrow}\left(1 - \frac{1-F(q)}{e^y}\right)\right) - F(q)}{1 - F(q)} = 1 - e^{-y}. \quad (5)$$

Conversely, assume that  $\eta_q$  is stochastically smaller than a standard exponential random variable, that is,  $P(\eta_q \leq y) \geq 1 - e^{-y}$  for all  $y \geq 0$ . With (4) we get that

$$\begin{aligned} \frac{G\left(F^{\leftarrow}\left(1 - \frac{1-F(q)}{e^y}\right)\right) - G(q)}{1 - G(q)} &\geq 1 - e^{-y} \iff \frac{1 - G\left(F^{\leftarrow}\left(1 - \frac{1-F(q)}{e^y}\right)\right)}{1 - G(q)} \leq e^{-y} \\ &\iff G\left(F^{\leftarrow}\left(1 - \frac{1 - F(q)}{e^y}\right)\right) \leq 1 - \frac{1 - G(q)}{e^y} \iff \\ &F^{\leftarrow}\left(1 - \frac{1 - F(q)}{e^y}\right) \leq G^{\leftarrow}\left(1 - \frac{1 - G(q)}{e^y}\right). \end{aligned}$$

Denote  $z_F = F^{\leftarrow}(1 - e^{-y}(1 - F(q)))$  and  $z_G = G^{\leftarrow}(1 - e^{-y}(1 - G(q)))$ . Since  $F(z_F) = 1 - e^{-y}(1 - F(q))$  and  $G(z_G) = 1 - e^{-y}(1 - G(q))$ , we have,

$$e^{-y} = \frac{1 - G(z_G)}{1 - G(q)} = \frac{1 - F(z_F)}{1 - F(q)}.$$

Further, since  $z_F \leq z_G$  then

$$\frac{1 - F(z_F)}{1 - F(q)} = \frac{1 - G(z_G)}{1 - G(q)} \leq \frac{1 - G(z_F)}{1 - G(q)}.$$

This observation completes the proof since  $z_F \in [q, \infty)$ . The proof of the second assertion is similar.

(iii) We have,

$$\frac{G(x) - G(q)}{1 - G(q)} \geq \frac{F(x) - F(q)}{1 - F(q)} \quad \forall x > q \geq x_0 \iff \frac{1 - G(x)}{1 - G(q)} \leq \frac{1 - F(x)}{1 - F(q)} \quad \forall x > q \geq x_0 \iff$$

$$\frac{1 - G(x)}{1 - F(x)} \leq \frac{1 - G(q)}{1 - F(q)} \quad \forall x > q \geq x_0 \iff \frac{1 - G(x)}{1 - F(x)} \text{ is nonincreasing for all } x > x_0. \blacksquare$$

### 2.3 Proof of Theorem 1.

Under the conditions of Theorem 1,  $F_0(X_1)$  is uniformly distributed on  $[0, 1]$ , that is,  $F_0(X_1) \sim U[0, 1]$ , hence  $-\ln(1 - F_0(X))$  is standard exponential random variable. It follows from Rényi's representation (see [4]), that

$$\left\{-\ln(1 - F_0(X_{(n-i)})) + \ln(1 - F_0(X_{(n-k)}))\right\}_{i=0}^{k-1} \stackrel{d}{=} \left\{\sum_{j=i+1}^k \frac{E_{n-j+1}}{j}\right\}_{i=0}^{k-1},$$

where  $E_1, E_2, \dots$  are independent standard exponential variables. Therefore the distribution of the left-hand side does not depend on  $n$  and

$$\{-\ln(1 - F_0(X_{(n-i)})) + \ln(1 - F_0(X_{(n-k)}))\}_{i=0}^{k-1} \stackrel{d}{=} \{E_{(k-i)}\}_{i=0}^{k-1},$$

where  $E_{(1)} \leq \dots \leq E_{(k)}$  are the  $n$ th order statistics of the sample  $\{E_i\}_{i=1}^k$ . Finally we have,

$$\sqrt{k}(R_{k,n} - 1) \stackrel{d}{=} \sqrt{k} \left( \frac{1}{k} \sum_{i=0}^{k-1} E_{(k-i)} - 1 \right) = \sqrt{k} \left( \frac{1}{k} \sum_{j=1}^k E_j - 1 \right),$$

and the assertion follows from the Central Limit Theorem.

## 2.4 Proof of Theorem 2.

We first prove (i). The steps of the proof are similar to corresponding steps in [6] and [7]. Consider asymptotic behavior of  $R_{k_n,n}$  as  $n \rightarrow \infty$ . Denote

$$Y_i = \ln(1 - F_0(q)) - \ln(1 - F_0(X_i^*)),$$

where  $\{X_i^*\}_{i=1}^{k_n}$  are i.i.d. random variables introduced in Lemma 1 with the distribution function

$$F_q(x) = \frac{F_1(x) - F_1(q)}{1 - F_1(q)}, \quad q < x.$$

Taking  $F = F_0$  and  $G = F_1$  we have,  $Y_i \stackrel{d}{=} \eta_q$ ,  $i \in \{1, \dots, k_n\}$ . Notice that, in view of Lemma 1, the joint distribution of order statistics  $\{Y_{(i)}\}_{i=1}^{k_n}$  of the sample  $\{Y_j\}_{j=1}^{k_n}$  is equal to the joint conditional distribution of order statistics  $\{Z_{(j)}\}_{j=1}^{k_n}$  of  $\{Z_j\}_{j=1}^{k_n}$  given  $X_{(n-k_n)} = q$ , where

$$Z_j = \ln(1 - F_0(X_{(n-k_n)})) - \ln(1 - F_0(X_{(n-j+1)})), \quad j = 1, \dots, k_n.$$

Clear,

$$R_{n,k_n} = \frac{1}{k_n} \sum_{i=1}^{k_n} Z_i.$$

So, the conditional distribution of  $R_{k_n,n}$  given  $X_{(n-k)} = q$  is equal to the distribution of  $\frac{1}{k_n} \sum_{i=1}^{k_n} Y_i$ . Further, distribution functions  $F_1$  and  $F_0$  satisfy  $B(F_0, F_1)$  or  $B(F_1, F_0)$ . First suppose that the condition  $B(F_0, F_1)$  holds for some  $\varepsilon > 0$  and  $x_0$ . Since  $x^* = +\infty$ ,  $X_{(n-k_n)} \rightarrow +\infty$  a.s., we may consider the case  $q > x_0$  only. Proposition 1 (iii) implies, that

$$\frac{1 - F_1(x)}{1 - F_1(x_0)} \geq \frac{(1 - F_0(x))^{1-\varepsilon}}{(1 - F_0(x_0))^{1-\varepsilon}}, \quad x > x_0.$$

With (5), we get that,

$$P(Y_1 \leq x) = 1 - \frac{1 - F_1 \left( F_0^{\leftarrow} \left( 1 - \frac{1 - F_0(q)}{e^x} \right) \right)}{1 - F_1(q)} \leq$$

$$1 - \frac{\left( 1 - F_0 \left( F_0^{\leftarrow} \left( 1 - \frac{1 - F_0(q)}{e^x} \right) \right) \right)^{1-\varepsilon}}{(1 - F_0(q))^{1-\varepsilon}} = 1 - e^{-(1-\varepsilon)x},$$

hence  $Y_1$  is stochastically larger than a random variable  $E \sim \text{Exp}(1 - \varepsilon)$ , write  $Y_1 \gg E$ . Further, let  $E_1, \dots, E_{k_n}$  are i.i.d. random variables with distribution function  $H(x) = 1 - e^{-(1-\varepsilon)x}$ , then

$$\sqrt{k_n} \left( \frac{1}{k_n} \sum_{i=1}^{k_n} Y_i - 1 \right) \gg \sqrt{k_n} \left( \frac{1}{k_n} \sum_{i=1}^{k_n} E_i - 1 \right). \quad (6)$$

Since (6) holds for all  $q > x_0$ , and  $X_{(n-k_n)} \rightarrow +\infty$  a.s. as  $n \rightarrow \infty$ , we have under the conditions of Theorem 2, that

$$\sqrt{k_n}(R_{k_n,n} - 1) \gg \sqrt{k_n} \left( \frac{1}{k_n} \sum_{i=1}^{k_n} E_i - 1 \right). \quad (7)$$

It follows from Lindeberg-Feller theorem, that

$$(1 - \varepsilon) \sqrt{k_n} \left( \frac{1}{k_n} \sum_{i=1}^{k_n} E_i - \frac{1}{1 - \varepsilon} \right) \xrightarrow{d} \xi \sim N(0, 1), \quad n \rightarrow \infty,$$

therefore

$$\sqrt{k_n} \left( \frac{1}{k_n} \sum_{i=1}^{k_n} E_i - 1 \right) \xrightarrow{P} +\infty, \quad n \rightarrow \infty. \quad (8)$$

Finally, with (7), we have,

$$\sqrt{k_n}(R_{k_n,n} - 1) \xrightarrow{P} +\infty, \quad n \rightarrow \infty.$$

If the condition  $B(F_0, F_1)$  holds, then

$$\sqrt{k_n}(R_{k_n,n} - 1) \xrightarrow{P} -\infty, \quad n \rightarrow \infty,$$

and the proof is similar. The second assertion easily follows from (7) and (8).

## 2.5 Proof of Theorem 3.

Firstly we prove (i). Denote  $\overline{F}(x) = 1 - F(x)$ . In notation of the proof of Theorem 2, find the distribution of  $Y_1$ . First assume that  $C(F_0, F_1)$  holds. With (5) and Proposition 1 (iii) we have,

$$P(Y_1 \leq x) = 1 - \frac{\overline{F}_1(\overline{F}_0^{\leftarrow}(\overline{F}_0(q)e^{-x}))}{\overline{F}_1(q)} \leq$$

$$1 - \frac{\overline{F}_0(q)e^{-x}(-\ln(\overline{F}_0(q)e^{-x}))^\varepsilon}{\overline{F}(q)(-\ln \overline{F}_0(q))^\varepsilon} = 1 - e^{-x} \left(1 + \frac{x}{-\ln \overline{F}_0(q)}\right)^\varepsilon.$$

For  $\varepsilon, c \in (0, 1)$ ,

$$(1 + cx)^\varepsilon \geq 1 + c\varepsilon - c\varepsilon e^{-x}, \quad x \geq 0,$$

and  $G(x) = 1 - e^{-x}(1 + c\varepsilon - c\varepsilon e^{-x})$  is the distribution function. Hence,

$$P(Y_1 \leq x) \geq 1 - e^{-x} - \frac{\varepsilon}{-\ln \overline{F}_0(q)}(1 - e^{-x}).$$

Further, let  $\zeta, \zeta_1, \dots, \zeta_{k_n}$  be i.i.d. random variables with this distribution function. Therefore, like the proof of Theorem 2,

$$\sqrt{k_n} \left( \frac{1}{k_n} \sum_{i=1}^{k_n} Y_i - 1 \right) \gg \sqrt{k_n} \left( \frac{1}{k_n} \sum_{i=1}^{k_n} \zeta_i - 1 \right). \quad (9)$$

Clear,

$$E\zeta = 1 + \frac{\varepsilon}{-2 \ln \overline{F}_0(q)}, \quad Var \zeta = 1 - \left( \frac{\varepsilon}{-2 \ln \overline{F}_0(q)} \right)^2,$$

so we have,

$$\sqrt{k_n} \left( \frac{1}{k_n} \sum_{i=1}^{k_n} \zeta_i - 1 \right) = \sqrt{k_n} \left( \frac{1}{k_n} \sum_{i=1}^{k_n} \zeta_i - E\zeta \right) + \sqrt{k_n} \frac{\varepsilon}{-2 \ln \overline{F}_0(q)}. \quad (10)$$

Consider now the statistic  $\sqrt{k_n}/\ln \overline{F}_0(X_{(n-k_n)})$ . Denote  $R_i = \overline{F}_1(X_i)$ ,  $i = 1, \dots, n$ . Since  $\overline{F}_1$  is continuous,  $R_1, \dots, R_n$  are i.i.d. standard uniform random variables and  $R_{(k_n)} = \overline{F}_1(X_{(n-k_n)})$ . Theorem 2.2.1 [4] implies, that

$$\frac{n}{\sqrt{k_n}} \left( R_{(k_n)} - \frac{k_n}{n} \right) \xrightarrow{d} N(0, 1), \quad n \rightarrow \infty. \quad (11)$$

Using the delta method (see [11]) for the function  $f(x) = -x/\ln x$ , we have

$$\frac{n}{\sqrt{k_n}} \left( \frac{R_{(k_n)}}{-\ln R_{(k_n)}} - \frac{k_n/n}{-\ln(k_n/n)} \right) \xrightarrow{P} 0, \quad n \rightarrow \infty,$$

since under the conditions of theorem

$$f' \left( \frac{k_n}{n} \right) = -\frac{1}{\ln(n/k_n)} + \frac{1}{(\ln(n/k_n))^2} \rightarrow 0, \quad n \rightarrow \infty.$$

Further,

$$\begin{aligned} & \frac{n}{\sqrt{k_n}} \left( \frac{R_{(k_n)}}{\ln R_{(k_n)}} - \frac{k_n/n}{\ln(k_n/n)} \right) = \\ & \frac{n}{\sqrt{k_n}} \left( \frac{R_{(k_n)}}{\ln R_{(k_n)}} - \frac{k_n/n}{\ln(R_{(k_n)})} \right) + \sqrt{k_n} \left( \frac{1}{\ln R_{(k_n)}} - \frac{1}{\ln(k_n/n)} \right), \end{aligned}$$



and (11) implies that the first summand in the right hand side tends to 0 in probability. Therefore,

$$\sqrt{k_n} \left( \frac{1}{\ln R_{(k_n)}} - \frac{1}{\ln(k_n/n)} \right) \xrightarrow{P} 0, \quad n \rightarrow \infty,$$

and under the conditions of Theorem 3,

$$\frac{\sqrt{k_n}}{-\ln R_{(k_n)}} = \sqrt{k_n} \left( \frac{1}{-\ln R_{(k_n)}} - \frac{1}{-\ln(k_n/n)} \right) + \frac{\sqrt{k_n}}{-\ln(k_n/n)} \xrightarrow{P} +\infty, \quad n \rightarrow \infty.$$

On the other hand, from (3) it follows that

$$\frac{\sqrt{k_n}}{-\ln \bar{F}_0(X_{(n-k_n)})} = \frac{\sqrt{k_n}}{-\ln \bar{F}_0(\bar{F}_1^{\leftarrow}(R_{(k_n)}))} \geq \frac{\sqrt{k_n}}{-\delta^{-1} \ln R_{(k_n)}} \xrightarrow{P} +\infty \quad (12)$$

as  $n \rightarrow \infty$ . Further, it follows from the Law of large numbers for triangular arrays (see [9]), that for any  $\epsilon > 0$

$$\sqrt{k_n} \left( \frac{1}{k_n} \sum_{i=1}^{k_n} \zeta_i - 1 \right) = o_P(k_n^\epsilon), \quad n \rightarrow \infty.$$

It means that the term in the left hand side is asymptotically smaller in probability than  $k_n^\epsilon$ . Hence for any  $q$ , given  $X_{(n-k_n)} = q$

$$\sqrt{k_n} \left( \frac{1}{k_n} \sum_{i=1}^{k_n} Y_i - 1 \right) \xrightarrow{P} +\infty, \quad n \rightarrow \infty,$$

and finally,

$$\sqrt{k_n} (R_{k_n, n} - 1) \xrightarrow{P} +\infty, \quad n \rightarrow \infty.$$

If the condition  $C(F_1, F_0)$  holds, then

$$\sqrt{k_n} (R_{k_n, n} - 1) \xrightarrow{P} -\infty, \quad n \rightarrow \infty,$$

and the proof is the same. The second assertion clearly follows from (9), (10) and (12).

## References

- [1] *Dufour R., Maag U.R.* Distribution Results for Modified Kolmogorov-Smirnov Statistics for Truncated or Censored Samples. — *Technometrics*, 1978, v. 20, p. 29–32.
- [2] *Guilbaud O.* Exact Kolmogorov-Type Test for Left-Truncated and/or Right-Censored Data. — *Journal of American Statistical Association*, 1998, v. 83, p. 213–221.

- [3] *Chernobai A., Menn C., Rachev S. T., Truck S.* Estimation of operational value-at-risk in the presence of minimum collection thresholds. — Tech. Rep., University of California, Santa Barbara, Calif, USA, 2005.
- [4] *Ferreira A., Haan L. de.* Extreme value theory. An introduction. N. Y.: Springer, Springer Series in Operations Research and Financial Engineering, 2006.
- [5] *Gardes L., Girard S., Guillou A.* Weibull tail-distributions revisited: a new look at some tail estimators. — Journal of Statistical Planning and Inference, 2009, v. 141, p. 429–444.
- [6] *Rodionov I. V.* A discrimination test for tails of Weibull-like distributions. — to appear in Probability Theory and its Applications.
- [7] *Rodionov I. V.* Discrimination of close hypotheses on distribution tails using higher order statistics. — to appear in Extremes.
- [8] *Haan L. de, Resnick S.* Second-order regular variation and rates of convergence in extreme value theory. — The Annals of Probability, 1996, v. 24, i. 1, p. 97–124.
- [9] *Mikusheva A. E.* The law of large numbers and the logarithmic law for arrays. — Fundamental and applied mathematics, 2000, v. 6, i. 1, p. 195–206 (in Russian).
- [10] *Falk M.* Some best estimators for distributions with finite endpoint. — Statistics, 1995, v. 27, p. 115–125.
- [11] *Resnick S. I.* A probability path. Birkhauser, 1999.